DOCUMENT RESUME

ED 076 146 HE 004 124

AUTHOR Aleamoni, Lawrence M.; And Others
TITLE Teacher Folklore and the Sensitivity of a Course Evaluation Questionnaire (Revised).
INSTITUTION Illinois Univ., Urbana. Office of Instructional Resources.
REPORT NO RR-313
PUB DATE Jan 72
NOTE 21p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Bibliographies; *College Faculty; *Course Evaluation; *Faculty Evaluation; *Higher Education; Research; Research Projects; *Student Attitudes

ABSTRACT

The purpose of this study was to determine if a) "Folklore" about a teacher contributes to his ratings on a course evaluation questionnaire and b) changes in students' attitudes during the course of instruction can be determined with a course evaluation questionnaire. Multivariate techniques and discriminate analysis were employed. The results indicated that there were no significant differences in attitudes towards the course in educational statistics between those who took the course in 1967-68 and those who took it in 1968-69. This seems to indicate that students do not build a "folklore" about a course based upon the course presented a year earlier. The results also indicated that changes in attitude about a course while the students are enrolled in that course can be measured by a course evaluation questionnaire. A 16-item bibliography is included. (Author/MJM)

FILMED FROM BEST AVAILABLE COPY

ED 076146

research report # 313

research report

Teacher Folklore and the Sensitivity of a Course Evaluation Questionnaire

(Revised)

Lawrence M. Aleamoni
Makonnen Yimer
J. Maurice Mahan

Measurement and Research Division
Office of Instructional Resources
307 Engineering Hall
University of Illinois

January, 1972

measurement and research division
office of instructional resources
university of illinois
urbana, illinois

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

ABSTRACT

The purpose of this study was to determine if (a) "Folklore" about a teacher contributes to his ratings on a course evaluation questionnaire and (b) changes in students' attitudes during the course of instruction can be determined with a course evaluation questionnaire. Multivariate techniques such as MANOVA and discriminate analysis are ideally suited for this type of research and were employed. The results indicated that there were no significant differences in attitudes towards the course in educational statistics between those who took the course in 1967-1968 and those who took it in 1968-1969. This seems to indicate that students' do not build a "Folklore" about a course based upon the course presented a year earlier. The results also indicated that changes in attitude about a course while the students are enrolled in the course can be measured by a course evaluation questionnaire.

TEACHER FOLKLORE AND THE SENSITIVITY OF A COURSE EVALUATION QUESTIONNAIRE

Lawrence M. Aleamoni, Makonnen Yimer, and J. Maurice Maham

University of Illinois

In an effort to improve the quality of instruction at all levels of education, many evaluation procedures and instruments have been developed. These procedures and instruments are usually designed to give feedback to the teacher so that he can take some action to improve his teaching and the subsequent performance of his students.

One of the methods of providing evaluative feedback is to measure the attitudes of students toward the teacher and the course. The authors have reviewed several of the instruments developed and used by various other universities (Coffman, 1954; Cosgrove, 1959; Isaacson, McKeachie, Milholland, Lin, Hofeller, Baerwaldt, and Zinn, 1964; Rees, 1969; Remmers and Elliott, 1949; and Yong and Sassenrath, 1969). The usual procedure for developing those questionnaires is that a group of items is constructed, given to students, factor analyzed, and revised. Items are retained which meet certain criteria. Finally norms are devised so that the teachers can compare their rating with other teachers. Often this is the end of the process, except for occasional renorming of the data.

It would seem that if an attitude questionnaire of this nature is to be useful to a teacher, a research program should be conducted along with it to determine what things affect the separate factors in the instrument as well as to determine its sensitivity to attitudes. It is possible that a given instrument does not actually measure the attitudes for which it was designed.

When students select courses, particularly at the graduate level, they usually talk to other students and professors about available courses and the teachers who teach them. They try to find out something about the content of the course, the text used, the projects required, and the teaching style of the instructor. Through these contacts it would seem reasonable to assume that students develop some attitudes about the course and its instructor before they go to class. (These attitudes may be favorable or unfavorable and of variable strength.)

If an instructor does a particularly good or bad job of teaching a course and this fact is passed on to other students, it would seem that if he taught the course again his new students' attitudes could be different from the incoming attitudes of the students of the previous year. Regardless of what the instructor may do in the class, the preconceived attitudes of the students could have an effect on their initial as well as on their final evaluation of the course during the second year.

The present study was designed to investigate the extent to which (a) "Folklore" about a teacher contributes to his ratings on a course evaluation questionnaire and (b) changes in attitude during the course of instruction can be determined with a course evaluation questionnaire.

Method

Subjects

The subjects (Ss) used in this study were two groups of graduate students who were enrolled in a graduate-level educational statistics course during the academic years 1967-1968 and 1968-1969 (see Table 1) taught by the same instructor.

TABLE 1

Number of Ss in each Classification

Year	Time	
	Pre	Post
1967 - 1968	21	21
1968 - 1969	24	24

Materials

The questionnaire used to collect student attitudes was the Illinois Course Evaluation Questionnaire (CEQ). The CEQ was developed to "elicit student opinions about a standardized set of statements relative to certain standardized aspects of an instructional program" (Spencer and Aleamoni, 1969). The CEQ consists of fifty items. The reliability of the total test has been calculated as .93 (using a Spearman-Brown correlation corrected for length) (Spencer and Aleamoni, 1970) and .93 using Cronbach's α on more recent data. The fifty items of the CEQ are grouped into six subscores (Table 2). Five of the subscores were developed by factor analysis and the sixth consists of items that did not load highly on the other factors but were retained because of their special interest to faculty members.

The product moment correlations between the subscores usually range from .46 to .77, while their reliability ranges from .80 to .98. The CEQ is printed on machine-scorable answer sheets. There are four response positions for each question which are: strongly agree, agree, disagree, and strongly disagree. The items are either stated negatively or positively. For positive statements a weight of 4, 3, 2, and 1 is assigned respectively for the response position, while for negatively stated statements the reverse is true.

TABLE 2

Definition of the Six Dependent Variates

Variate No.	Subscore	No. of Items
1	General Course Attitude	8
2	Method of Instruction	8
3	Course Content	8
4	Interest and Attention	8
5	Instructor	8
6	Specific Items	10

Design

The specific hypotheses to be tested were:

1. That there would be no significant differences in the evaluations of the course between the first and second time the instructor taught the course (year).
2. That there would be no significant differences between the evaluations which were collected at the beginning and the end of the course (time).
3. That there would be no significant interaction between the evaluation of the course when the instructor taught the first year and the second year and the evaluation of the course at the beginning and end of a semester.

The dependent variables were ratings of the six subscores of the questionnaire. The appropriate procedure of analysis for this design was a 2 by 2 multivariate analysis of variance (MANOVA) with six dependent variables. The

Ss were nested under the year variable and there were repeated measures on the time variable. A discriminant analysis was done for the main and interaction effects.

Procedure

The CEQ was administered to students in both groups (the 1967-68 and the 1968-69 class) both at the beginning and end of the course. Each group was also informed at the time of administration about the second administration of the same instrument at the end of the semester. Anonymity and identification of a subject's response was made possible by the fact that each individual used an arbitrary number unknown to the instructor.

The data of 4 Ss from the first group and 3 Ss from the second group were excluded from the analyses because either they did not make both pre and post evaluations of the course or did not complete the questionnaire. The probability level adopted for significance testing was .05.

Results

Before a MANOVA was done, a correlation coefficient over item mean response for each subscore and for the total subscore was determined for YEAR and TIME. This was done to determine the relationships of the responses between the two groups of students as well as within the same group. The correlations are presented in Table 3. Under "YEAR" the correlations between the beginning of the semester evaluation of 1967-1968 with that of 1968-69, described as PRE-PRE, and the correlation between the evaluation at the end of the semester of 1967-1968 with that of 1968-1969, described as POST - POST are given. For instance, the correlation of 0.78 for General Attitude under "YEAR PRE - PRE" was arrived at by correlating the item mean responses of eight items between the two initial evaluations of the course. In every respect, there seems to

TABLE 3

Correlation By YEAR and TIME

VARIATES	YEAR			TIME	
	(1967-68, 1968-69) PRE - PRE	(1967-68, 1968-69) POST - POST	(1967-68, 1968-69) POST - PRE	(1967-68) PRE-POST	(1968-69) PRE-POST
General Course Attitude 01	.78	.84	.63	.61	.84
Method of Instruction 02	.05	-.15	-.59	.29	.44
Course Content 03	.92	.99	.89	.78	.89
Interest and Attention 04	.93	.89	.78	.64	.82
Instructor 05	.67	.82	.30	.81	.50
Specific Items 06	.88	.83	.61	.32	.79
Total (over all)	.82	.90	.60	.51	.69

be a high positive agreement in the evaluation of the two groups except Variable 2 (Method of Instruction) with practically no agreement. From this it may be observed that (1) the students held generally the same attitude during the two course offerings as the PRE - PRE correlation indicates; and (2) the instructor was consistent in conducting the courses bringing about the same effect as the POST - POST correlation shows. The low correlation of Variable 2 merely shows that there is no relationship between the evaluation of the individuals with respect to this dependent variable. This could mean that the instructor changed his method of instruction from one year to the next or that the students information sources were not reliable.

Table 3 also shows the correlation between the beginning and end of the students' evaluation for each semester with respect to time. The lower correlation compared to the YEAR shows that there is more disagreement in the evaluation between the beginning and end of a course during a semester than between the years. This is clearly shown by the overall correlation of .57 for 1967-1968 and .69 for 1968-1969 while the overall correlation of the PRE - PRE and POST - POST was .62 and .90 respectively.

These correlations seem to suggest that there is not much difference in the evaluations between the two years while there is a much larger difference between the evaluations within the same year.

The means and standard deviations for each subscore and the mean and standard deviations of item responses for each subscore are presented in Table 4. The mean item responses are included for purposes of interpretation and indicate the mean response given for the items within each subscore.

TABLE 4

Means and Standard Deviations for the
Six Dependent Variables.

Year	Variate	PRE				POST			
		Subscore		Item		Subscore		Item	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1967	01	13.10	2.54	1.64	.24	16.29	4.13	2.04	.23
	02	14.43	2.06	1.80	.23	22.57	5.10	2.82	.14
	03	17.05	2.46	2.13	.43	19.14	3.17	2.39	.72
	04	15.81	3.57	1.98	.30	17.81	4.56	2.23	.16
	05	13.57	3.25	1.70	.19	16.52	4.45	2.07	.55
	06	19.43	2.48	1.94	.27	24.33	3.58	2.43	.36
1968	01	14.20	3.71	1.78	.14	16.46	3.80	2.06	.20
	02	15.96	2.84	2.00	.19	23.46	3.52	2.93	.08
	03	17.21	2.61	2.15	.37	19.75	2.65	2.47	.65
	04	15.71	3.30	1.97	.23	19.00	4.20	2.38	.24
	05	14.63	2.58	1.83	.14	17.33	2.67	2.17	.47
	06	19.83	3.21	1.98	.23	22.75	2.59	2.28	.40

Tables 5 and 6 contain the within cell correlations and standard deviations between variates for the time and year effect. The mean products matrices for between years, time, and interactions are shown in Table 7. These matrices were used in the computation of the significant tests. Table 8 indicates the degrees of freedom used in the MANOVA.

TABLE 5

Within - Cell Correlations

Between Variates for Year Effect

VARIATES	01	02	03	04	05	06
01	(4.429)*					
02	.57	(3.454)				
03	.74	.60	(3.153)			
04	.67	.49	.68	(4.737)		
05	.38	.39	.13	.23	(3.445)	
06	.55	.44	.58	.28	.52	(3.066)

*Within-cell standard deviations appear as diagonal entries.

TABLE 6

Within - Cell Correlations

Between Variates for Time Effect

VARIATES	01	02	03	04	05	06
01	(2.755)*					
02	.51	(3.773)				
03	.52	.48	(2.375)			
04	.76	.52	.54	(3.142)		
05	.85	.61	.47	.58	(3.260)	
06	.60	.66	.55	.54	.64	(3.056)

*Within-cell standard deviations appear as diagonal entries.

TABLE 7

Mean Products

SOURCE	VARIATES	01	02	03	04	05	06
Between Years	01	9.26					
	02	17.40	32.70				
	03	5.53	10.39	3.30			
	04	7.84	14.74	4.68	6.64		
	05	13.41	25.21	8.01	11.37	19.44	
	06	-8.49	-15.95	-5.07	-7.19	-12.30	7.78
Between Time	01	162.68					
	02	471.90	1368.90				
	03	141.17	409.50	122.50			
	04	162.68	471.90	141.71	162.68		
	05	170.74	495.30	148.17	170.74	179.21	
	06	232.59	674.70	201.83	232.59	244.12	332.54
Interaction Year & Time	01	4.95					
	02	3.39	2.30				
	03	-2.36	-1.61	1.12			
	04	-6.80	-4.65	3.23	9.35		
	05	1.25	.88	-.61	-1.77	.33	
	06	10.47	7.16	-4.96	-14.38	2.72	22.14

TABLE 8

Source Table for the MANOVA

Sources	df
Year	1
error (a)	43
Time	1
Year x Time	1
error (b)	43
Total	89

Since the degrees of freedom of each hypothesis (the Time, Year, and Interaction effect) is one, the three significance tests, (a) the likelihood ratio F test, (b) the Trace T, and (c) Roy's criterion, are equivalent (Jones, 1966). This means that the significance level for one of the tests applies equally well for the other two.

Table 9 contains the tests of significance for the Time, Year, and Interaction effects. For the Time effect the MANOVA F is 16.6155, and Trace T is 2.6235, and Roy's criterion is 0.7240. These three tests are highly significant with a probability of less than 0.005. For the Year effect the MANOVA F is 2.0214, the Trace T is 0.3042, and Roy's criterion is 0.2419. These values are not considered significant ($p > .09$). The values for the significance of the interaction effect are F = 2.2918, Trace T = 0.3622, and Roy's criterion = 0.2657. These values are not considered significant ($p > .06$).

TABLE 9

Tests of Significance

DEPENDENT VARIABLES = 6	EFFECT		
	Time	Year	Interaction
F	16.6155	2.0214	2.2918
df (numerator)	6	6	6
df (denominator)	38	38	38
Probability	$p < .005$	$p > .09$	$p > .06$
Trace T	2.6235	.3042	.3622
Tabled Proficiency Roy's Criterion	.7240	.2419	.2657

Discriminant function for the Time effect.

V_T (normalized) = $.554x_1 + .634x_2 + .048x_3 - .246x_4 - .478x_5 - .004x_6$

V_T (standardized) = $2.454x_1 + 2.190x_2 + .151x_3 - 1.165x_4 - 1.647x_5 - .012x_6$

In an effort to determine the nature of the difference in attitude between the beginning and end of the course, a discriminant analysis was computed for the data. Discriminant analysis is usually used to discriminate between two or more groups of subjects (Rao, 1952; Cooley and Lohnes, 1962; Jones and Bock, 1960; Tatsuoka and Tiedemann, 1954). It can also be used for "a more basic scientific purpose than taxonomic decisions, by revealing the dimensions along which several groups differ from one another" (Tatsuoka, 1969). The use of the discriminant analysis in this way determines the "linear combination of variables most sensitive to departure from the null hypothesis, in the sense that the sum of squares for hypothesis for the combination is a maximum with respect to

sum of squares for error" (Bock and Haggard, 1969). The standardized discriminant coefficients are determined by multiplying the raw discriminant function weights by the within cell standard deviations of the respective dependent variables (Jones, 1966; Tatsuoka, 1970). (Since there was only one degree of freedom for each hypothesis, there can only be one discriminant function for each effect.) The normalized and standardized discriminant functions for the time effect are presented in Table 9. The mean discriminant scores for the PRE and POST measurements for the two years are presented in Table 10.

TABLE 10

Mean Discriminant Scores

Using the Normalized Equation for Time Effect

1967 - 1968	1968 - 1969
$V_{PRE} = 6.640$	$V_{PRE} = 7.875$
$V_{POST} = 11.878$	$V_{POST} = 11.892$

Discriminant functions for the year and interaction effect are not reported since neither were significant at the .05 level.

Discussion

Since the interaction effect was not significant, the null hypothesis that there is no significant interaction between the course evaluations over the two years and the course evaluations collected at the beginning and end of the semester was supported. This essentially means that the attitude change due to year and time effect was additive.

In considering the non-significant F ratio for the year effect and the mean item responses, it would seem that conditions in this course were stable in terms of the effect on the students. The contention that a "folklore" which would affect the ratings would build up about the instructor after the first year he taught the course was not supported by the MANOVA results or the results of the POST - 1967 and PRE - 1968 correlation. However, "folklore" might be defined by students carrying over impressions from other courses taught, being optimistic, and thinking that things are going to be better the following year, thereby, rating the course and instructor much higher than they would at the end of the course.

In every case, however little the change in the attitude score for year effect, the change was generally positive except for variable 4 (Interest and Attention) in PRE and variable 6 (Specific Items) in POST.

The Time effect results indicate that there was a significant difference for the mean vectors of the six dependent variables between the beginning and end of the course for both years. It should also be noted here that the attitude change is in the positive direction. This appears to be a stable result in that approximately the same result occurs both years. Table 11 also shows that the time effect alone does not completely discriminate student attitude change about the course. In fact, the discriminatory power (Tatsuoka, 1970) of the time discriminant function is about 41%. This implies that the total variability of the attitude of students is not explainable by the time factor alone with respect to the dependent variables. It also implies that the dependent variables either do not quite measure what they are supposed to measure or additional measures are needed to clearly discriminate the attitude change.

TABLE 11

Distribution of Discriminant Scores

For Time Effect

Discriminant Scores	Time Effect	
	PRE	POST
16.50 - 18.49		11111
14.50 - 16.49		111111
12.50 - 14.49	1	11111111
10.50 - 12.49	1111	111111111111
8.50 - 10.49	111111	11111
6.50 - 8.49	1111111111111111	1111111
4.50 - 6.49	111111111111	11
2.50 - 4.49	11	
.50 - 2.49	1	
-1.50 - .49	1	

The standardized discriminant function for the time effect is: $V_T = 2.454x_1 + 2.190x_2 + .151x_3 - 1.165x_4 - 1.647x_5 - .012x_6$. This function indicates that subscores 1 (General Course Attitude), 2 (Method of Instruction), 5 (Instructor), and 4 (Interest and Attention) seem to be sensitive to any departure from the null hypothesis.

The standardized discriminant function shows that the group as described by the discriminant function, has a favorable general course outlook with a good attitude on the method of instruction. It also seems to show that students did have a positive attitude change toward the course during each semester, i.e., over the two year period. The within-cell correlations of the subscores

(Table 6) will add some light to the above discussion. The higher correlation of Variables 1 and 2 with Variables 4 and 5 compared to the correlation of Variable 1 with Variable 2, is evidence that General Course Attitude and Method of Instruction are more sensitive than the rest of the variables. In addition, Variables 1 and 2 contribute to the negative weighting of Variables 4 and 5. The most influential in the negative weighting in this respect, is the General Course Attitude since it correlates .76 and .85 with Interest-Attention (4) and Instructor (5), respectively. Hence, it is safe to say that General Course Attitude and Method of Instruction are the two main variables around which the attitude change occurred.

Variable 3 (Course Content) has a low positive weight. In addition, it has a low correlation with the rest of the dependent variables. Therefore, course content does not seem to contribute in discriminating student attitude between the beginning and end of the course offering. Variable 6 (Specific Items) is negatively weighted with practically zero weight.

Summary and Conclusions

The results indicate that there were no significant differences in attitudes toward a course in educational statistics between those who took the course in 1967-68 and those who took it in 1968-69. This seems to indicate that students' do not build a folklore about a course based upon the course presented a year earlier.

The results also indicate that changes in attitude about a course, as measured by the CEQ, can occur while students are enrolled in the course. It also indicates that the greatest change in attitude was in the areas of General Course Attitude and Method of Instruction. The Instructor as well as Interest-Attention scale were sensitive to attitude change in the negative direction,

but these variables are highly correlated with General Course Attitude and Method of Instruction which decreases their importance in their sensitivity to the attitude change. The foregoing indicates that changes in attitude during the course of instruction can be determined with a course evaluation questionnaire.

This study should serve to emphasize the need for research on attitude questionnaires to determine if each is a valid and useful instrument to measure what it was designed to measure. It appears that too often persons go through elaborate procedures to develop questionnaires and then do not take the time to do research on the nature of the instrument; to develop and determine its validity and usefulness. Perhaps there would be fewer non-significant results in the psychological and educational research literature if more research were done on the instruments which are used to collect data. Multivariate techniques such as MANOVA and discriminant analysis are ideally suited for this type of research and the availability of computers on which to process this data now makes it possible for more of this research to be undertaken.

References

Bock, R. D. & Haggard, E. A. The use of multivariate analysis of variance in behavioral research. In D. R. Whatla (Ed.), Handbook of measurement and assessment in behavioral sciences, reading. Mass.: Addison-Wesley, 1968.

Coffman, W. E. Determining students' concepts of effective teaching from their ratings of instructors. Journal of Educational Psychology, 1954, 45, 277-286.

Cooley, W. W. & Lohnes, P. R. Multivariate procedures for the behavioral sciences. New York: Wiley, 1962.

Cosgrove, D. J. Diagnostic rating of teacher performance. Journal of Educational Psychology, 1959, 50, 200-204.

Isaacson, R. L., McKeachie, N. G., Milholland, J. E., Lin, Y. G., Hofeller, M., Baerwaldt, J. W., and Zinn, K. L. Dimensions of student evaluations of teaching. Journal of Educational Psychology, 1964, 55, 344-351.

Jones, L. V. Analysis of variance in its multivariate developments. In R. B. Cattell (Ed.), Handbook of multivariate experimental psychology. Chicago: Rand McNally & Co., 1966.

Jones, L. V. & Bock, R. D. Multiple discriminant analysis applied to "Ways to live" ratings from six cultural groups. Sociometry, 1960, 23, 162-176.

Rao, C. R. Advanced statistical methods in biometric research. New York: Wiley, 1970

Rees, Richard D. Dimensions of students points of view in rating college teachers. Journal of Educational Psychology, 1969, 60, 476-482.

Remmers, H. H., and Elliot, D. N. The Indiana college and university staff evaluation program. School and Society, 1949, 70, 168-171.

Spencer, R. E. and Aleamoni, L. M. The Illinois Course Evaluation Questionnaire: a description of its development and a report of some of its results. Research Report No. 292, Urbana, Illinois: University of Illinois, Office of Instructional Resources, 1969.

Spencer, R. E. and Aleamoni, L. M. A student course evaluation questionnaire. Journal of Educational Measurement, 1970, 1, 209-210.

Tatsuoka, M. M. Multivariate analysis. Review of Educational Research, 1969, 39, 739-747.

Tatsuoka, M. M. and Tiedeman, D. V. Discriminate analysis. Review of Educational Research, 1954, 24, 402-420.

Tatsuoka, M. M., Discriminant analysis: the study of group differences. Champaign: IPAT, 1970.

Yong, G. D. and Sassenrath, J. M. Student personality correlates of teacher ratings. Journal of Educational Psychology, 1968, 59, 44-52.